Acta Cryst. (1999). A**55**, 305–313

Chemical bonding effects in the determination of protein structures by electron crystallography

Steven Chang,[a] Teresa Head-Gordon,[a] Robert M. Glaeser[a,b] and Kenneth H. Downing[a]*

[a]*Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, and* [b]*Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. E-mail: khdowning@lbl.gov*

(*Received 28 January 1998; accepted 16 July 1998*)

Abstract

Scattering of electrons is affected by the distribution of valence electrons that participate in chemical bonding and thus change the electrostatic shielding of the nucleus. This effect is particularly significant for low-angle scattering. Thus, while chemical bonding effects are difficult to measure with small-unit cell materials, they can be substantial in the study of proteins by electron crystallography. This work investigates the magnitude of chemical bonding effects for a representative collection of protein fragments and a model ligand for nucleotide-binding proteins within the resolution range generally used in determining protein structures by electron crystallography. Electrostatic potentials were calculated by *ab initio* methods for both the test molecules and for superpositions of their free atoms. Differences in scattering amplitudes can be well over 10% in the resolution range below 5 Å and are especially large in the case of ionized side chains and ligands. We conclude that the use of molecule-based scattering factors can provide a much more accurate representation of the low-resolution data obtained in electron crystallographic studies. The comparison of neutral and ionic structure factors at resolutions below 5 Å can also provide a sensitive determination of charge states, important for biological function, that is not accessible from X-ray crystallographic measurements.

1. Introduction

Electron crystallography is developing as a complement to X-ray crystallography in the determination of protein structures. The very strong scattering of electrons by matter makes it possible to work with extremely small amounts of material and indeed to use crystals no more than a single layer thick. Electron crystallography is thus particularly good for the study of monolayer crystals, such as those formed by integral membrane proteins, which have so frequently resisted attempts to form crystals suitable for X-ray work.

Because electrons are scattered by the (shielded) Coulomb potential, electron crystallography produces a map of the potential within the sample. X-ray crystallography, on the other hand, produces a map of the electron charge density. The electron density (ρ) and the Coulomb potential (φ) maps are related to each other by the Poisson equation,

$$\nabla^2\varphi = -4\pi\rho/\varepsilon, \tag{1}$$

where ε is the dielectric constant. The Coulomb potential and the electron charge density functions for isolated atoms are similar to one another in that they both are cusp-shaped functions, centered on the nucleus, with a width that approximates the size of the atom. Thus both the potential and the electron charge density give accurate descriptions of the atomic positions.

To date, atomic models of three proteins have been determined from electron crystallographic data. Although the resolution of the data was not high by the standards of current work in X-ray protein crystallography, the reliability of phases determined by analysis of high-resolution micrographs makes it possible to obtain a density map that can be interpreted with little ambiguity. A model of bacteriorhodopsin (bR) was constructed starting with a density map that had 3.5 Å resolution in the plane of the crystal and around 6 Å perpendicular (Henderson *et al.*, 1990). This model was subsequently refined to more isotropic resolution, with the inclusion of diffraction data that extended to higher tilt angles (Grigorieff *et al.*, 1996), and more recently an independent density map has been obtained at a resolution of 3.0 Å (Kimura *et al.*, 1997). A model for the green plant light harvesting complex (LHC) was based on a density map with nominal resolution of 3.4 Å (Kühlbrandt *et al.*, 1994). The structure of the tubulin $\alpha\beta$ dimer has recently been determined with a resolution of 3.7 Å (Nogales *et al.*, 1998). In all of these structures, the α-helix and β-strand segments were well resolved. Loop regions were not always as well defined in the density maps, but the resolution in each case was good enough to ensure that the correct topology was determined. In each of these cases though, electron diffraction intensities are available that extend to higher resolution than phases from the images, and extension of the map resolution would be beneficial in defining amino acid side-chain positions and interactions more clearly. This can be performed by refinement of the atomic model, following procedures that are routine in X-ray crystallography.

ISSN 0108-7673

In the course of building and refining an atomic model, the validity of the model is tested by computing *R* factors that relate the observed diffraction amplitudes, F_{obs}, to those calculated from the model, F_{calc},

$$R = \sum |F_{obs}(s) - F_{calc}(s)| \Big/ \sum F_{obs}(s), \qquad (2)$$

where

$$s = 2\sin(\theta/2)/\lambda, \qquad (3)$$

and the sums are over the set of diffraction amplitudes. In X-ray crystallography, it has been found empirically that an *R* factor below 20% gives a high level of confidence in the model, while *R* factors of greater than 30% may indicate serious misinterpretation of the map and therefore an incorrect structural model (Branden & Jones, 1990). In the structures determined by electron crystallography, initial *R* factors were found to be around 45%, and in the best case of refinement could be reduced to as low as 30% (Grigorieff *et al.*, 1996). Although these *R* factors are high by X-ray standards, the high quality of phases determined by electron crystallography produces maps that are clear enough that there can be little doubt that they have been interpreted without gross error. It is generally recognized that electron diffraction amplitudes are not as accurate as those currently obtained with proteins by X-ray diffraction (Wang & Kühlbrandt, 1992). Errors in the measurements will thus contribute to an *R* factor that is higher in electron crystallography. Here we address the point that the large values of *R* factors seen in the electron crystallographic structures may reflect errors in computing F_{calc} as well as inaccuracies in the model and the experimental data. More accurate calculations of F_{calc} would reduce the values of the *R* factors, thereby increasing our confidence in the details of the atomic model.

The outer region of the electrostatic potential of an atom is quite sensitive to changes in electron distribution that arise from formation of chemical bonds or ionic species. At high resolution (~1–2 Å), the scattering is dominated by penetration of the incident electron deep into the atomic core, where deviation of the shielded Coulomb potential from that in an isolated atom is negligible. However, at lower resolution the scattering is dominated by values of the Coulomb potential at points further from the shielded nucleus, where deformations of the electron distribution due to molecular bonding as well as the inherent asymmetry of the valence-electron orbitals can be quite pronounced. Thus, atomic spherically symmetric electron scattering factors are a poor approximation to the actual electron scattering factors at a resolution below 2.5 Å. While effects due to covalent bonding may produce important deviations, the effect of ionization on electron scattering is even more dramatic. Atomic scattering factors can be changed several-fold, and even reversed in sign, by atomic ionization (Vainshtein, 1964).

Bonding effects can play a significant part in electronic and mechanical properties of materials such as semiconductors and intermetallic alloys, and numerous attempts have been made to measure deviations from free-atom scattering in order to understand these properties using electron diffraction methods. Anstis *et al.* (1973), for example, demonstrated the strong effects that can make lattice image interpretation ambiguous in electron micrographs of a tungsten–niobium oxide. This work has nevertheless been made difficult in materials research by the fact that unit-cell sizes in most of the crystals are so small that at the lowest resolution for which data can be recorded, typically 2–5 Å, the chemical bonding effect is already very small. Fox (1993) addressed this problem in the context of experimental electron diffraction measurements, and Spence (1993) has given a thorough review of the problem in materials research. Numerous examples exist in the X-ray literature of calculation of charge distributions from diffraction data, and the methodology for this work is well developed (Su & Coppens, 1992). Koritsanszky *et al.* (1997) have recently shown examples of experimental determination of charge distribution in small organic molecules and discussed the technical problems and potentials of this approach in X-ray work.

For protein structure determination, the electron scattering cross section is expected to be significantly affected by chemical bonding over the entire resolution range for which data are typically collected. Ionization is an extreme case of these effects, and the strong distortions in map densities anticipated for ions were first observed in the LHC structure (Kühlbrandt *et al.*, 1994), where arginine and glutamic acid residues in particular display a strong increase and a strong decrease in density in the map, respectively, matching the expected positive and negative charges of each type of side chain. The effect with respect to ionic species is also apparent in the new density map of bR (Kimura *et al.*, 1997), where an attempt was made to judge ionization states of many of the functional side chains by differences in observed and expected densities in the electrostatic potential map. The densities of Asp85 and Asp212, for example, are low in the map computed with the full data set (60–3 Å), consistent with these two side chains being negatively charged as inferred from spectroscopic studies. In a novel and compelling analysis, Kimura *et al.* (1997) further show that the densities for both side chains increased substantially when scattering data in the range 3–7 Å, which are less affected by imperfect shielding, were used to calculate the map. This comparison demonstrates that the low density in maps computed from the full data is a result of the reversal in the sign of the scattering factor at low resolution (below ~5 Å), rather than being due to structural disorder of the side chains. These studies illustrate a strength that electron microscopy can have relative to X-ray crystallography: the combination of low- and high-resolution

data can be used to determine the charge state of amino acids important for biological function.

It is clear that values of F_{calc}, which are needed to calculate the *R* factor and guide the refinement process, could be improved by proper treatment of chemical bonding effects. The use of inaccurate calculated amplitudes is bound to increase the value of the *R* factor used to judge the progress towards a correct atomic model. Progress in refinement is judged by a decrease in the *R* factor as refinement proceeds. However, the possibility exists that structural errors in the calculated model would be introduced at the refinement stage in order to compensate for poor molecular structure factors when minimizing the *R* factor. The magnitude of the chemical bonding effects for covalently bonded materials (as opposed to ionic species) has not been studied in the context of protein electron crystallography, and thus we have had no prediction of how serious this problem could be. In the studies presented here, we have shown that there may be very significant alterations in structure factors caused by bonding effects, certainly large enough to produce a misleadingly high *R* factor.

This work investigates the magnitude of the chemical bonding effect for a representative collection of molecular fragments that are common in proteins, within the resolution range that includes most of the data used in determining protein structures by electron crystallography. We have also modeled the ribose triphosphate (RTP) moiety of adenosine triphosphate as a model for the ligand in nucleotide-binding proteins. The electrostatic potentials of these simple model compounds were computed based on both free-atom (spherical) scattering factors and on accurate molecular orbital calculations of the potential. Comparison of these potentials, and their Fourier transforms, demonstrates that errors in scattering amplitudes that are well over 10% for polar groups can be expected at resolutions below 5 Å, if scattering factors for neutral atoms are used to calculate the molecular structure factors.

2. Methods

We selected several types of molecules that represent common fragments found in proteins and for which we expected chemical bonding effects to be significant. Formamide was used as a model of the amide bond. Because hydrogen bonding has an important influence on the electron density between covalent fragments in a protein, we also evaluated the effect for a pair of formamide molecules forming an idealized hydrogen bond. We expect fully charged groups to show the greatest deviation from the free-atom approximation, of course, and the aspartic acid side chain was chosen as a representative of this class. Fully charged side chains are rarely found unpaired or unsolvated, so we also evaluated the more typical instance of ion pairs in proteins: we selected the Asp215 and Lys219 ion pair from thermolysin (3tln in the Protein Data Bank) for this purpose. Finally, we examined chemical bonding and ionization effects for the RTP fragment in the conformation in which it is a bound ligand in actin (1atn), as a function of net charge: neutral and −4.

All *ab initio* calculations were performed with the *Gaussian92* (*G92*; Frisch *et al.*, 1992) molecular orbital package. The geometries of each molecule for the electrostatic potential calculation were determined as follows. For the single formamide and the model of the aspartic acid side chain, we performed a full geometry optimization using HF/6-31G*. Based on the fully optimized formamide geometry, two formamide molecules were oriented in such a way as to produce an idealized hydrogen bond. For the ion pair and RTP, we used the atom positions from the PDB and optimized only the position of the H atoms using HF/6-31G*. Default convergence criteria defined in *G92* were used for all optimizations.

The electrostatic potential maps for both free atoms and for molecules were generated at the HF/6-31G* level of theory. To generate the numerical electrostatic potential maps for formamide and the aspartic acid side chain, we used a cubic grid with a side of 12.8 Å, with data points sampled every 0.2 Å. For the formamide pair, ion pair and RTP we used a rectangular grid with one side of 25.6 Å and the other two sides of 12.8 Å, with data points sampled every 0.2 Å. Potentials for the free-atom models were calculated as sums of the individual atom potentials, with the atom placed at the positions used in the determination of the molecular potential.

We found large differences in the electrostatic potential between molecule and free-atom calculations near the centers of the atoms. These differences are attributable to numerical errors in calculating the potential in regions where its gradient is extremely steep and have little to do with chemical bonding effects. Since we are interested in the electrostatic potential in the bonding region due primarily to valence electrons, data values within 0.3 Å from the atom centers were truncated and assigned the average value of the potential at a radius of 0.3 Å around each atom. We chose the value of 0.3 Å for the cutoff because the gradient at that radius is smaller, and the computed potential is thus more consistent between free atoms and molecules and yet still shows none of the anisotropy that develops in the potential at higher radius. To better describe the rapidly changing potential near the atom centers, the potential out to 2.0 Å from each atom center was recalculated using a finer sample density of 0.02 Å (0.05 Å for RTP) for both the molecule and free-atom models.

For the free-atom models, radially symmetric potentials for each of the atoms were calculated from the free-atom potentials generated by *G92*. We performed a linear interpolation between grid points to generate the

radial potential curve. Transforms of the radial potentials that we generated for free atoms were then compared with atomic scattering factors calculated by Doyle & Turner (1968) and Doyle & Cowley (1974).

Molecular structure factors were computed for the two versions of the three-dimensional potential function for each molecule, one version obtained from the molecular-orbital calculations of the Coulomb potential and the other from the superposition of independent atom potentials. R factors between the two versions were calculated within resolution zones as sums of the absolute value of the differences at each point in the transform divided by the average, summed over all points within the resolution zone,

$$R = \frac{\sum |F_{\text{molec}}(s) - F_{\text{free}}(s)|}{\sum [F_{\text{molec}}(s) + F_{\text{free}}(s)]/2}. \qquad (4)$$

3. Results

Fig. 1 shows the molecules used in the calculations. Figs. 2–5 show electrostatic potential maps, and their Fourier transforms, for the model compounds. Atom positions are as indicated in Fig. 1. For each figure, (*a*) shows a section of the three-dimensional potential calculated as the superposition of free atoms, (*b*) shows the corresponding section of the three-dimensional molecular-orbital calculation of potential, (*c*) shows the difference between free-atom and bonded-atom calculations [same section as (*a*) and (*b*)], (*d*) shows the

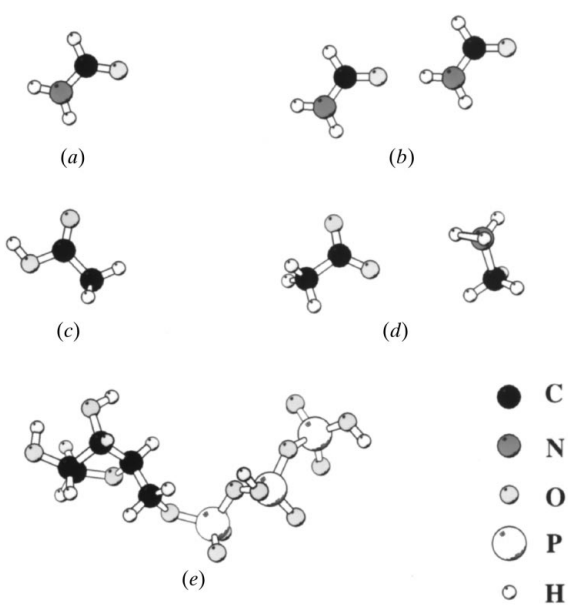

Fig. 1. Diagrams of the models used in computing chemical bonding effects, in the orientations shown in Figs. 2–5 and Fig. 8: (*a*) formamide, a model for an amide bond; (*b*) a pair of formamide molecules oriented to form an idealized hydrogen bond; (*c*) side chain of aspartic acid, expected to be negatively charged on the O atoms; (*d*) shortened versions of the Asp–Lys residues which form an ion pair in thermolysin; (*e*) ribose triphosphate.

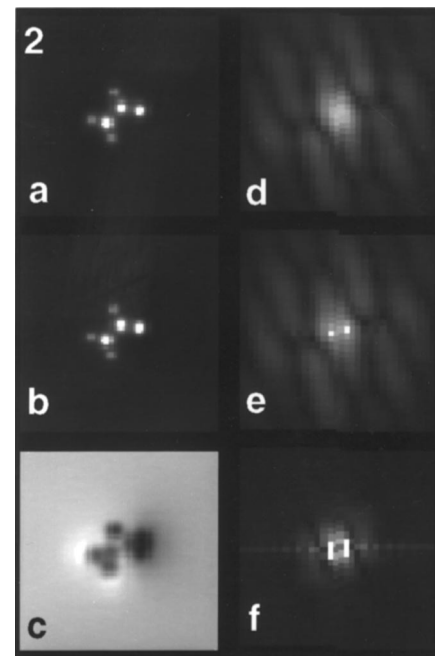

Fig. 2. Electrostatic potential map and its Fourier transform for formamide (amide). See text.

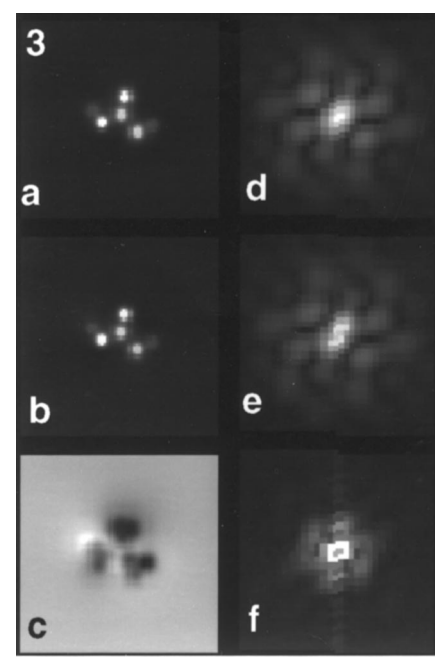

Fig. 3. As Fig. 2 but for aspartic acid. See text.

central section of the three-dimensional Fourier transform of the free-atom potential [the edge of this image corresponds to $s = 1.2$ in Table 1 (0.8 Å resolution)], (*e*) shows the central section of the transform of the molecular orbital potential, and (*f*) shows the central section of the three-dimensional transform of the difference between free- and bonded-atom potentials.

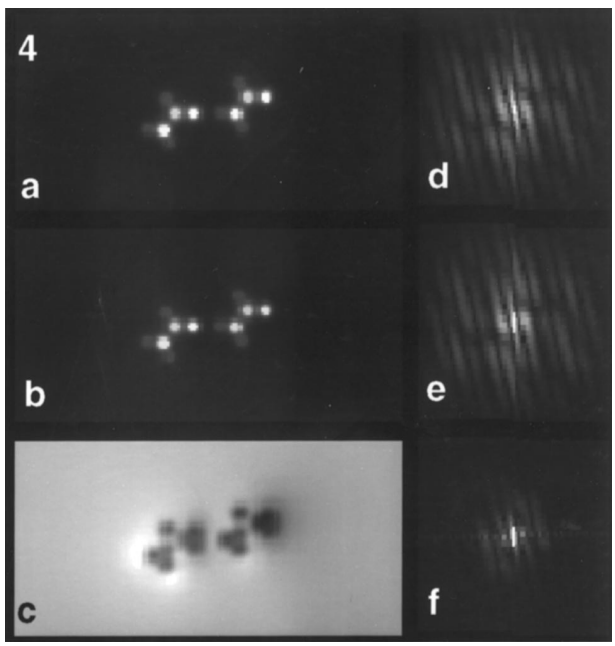

Fig. 4. As Fig. 2 but for a formamide pair (hydrogen bond). See text.

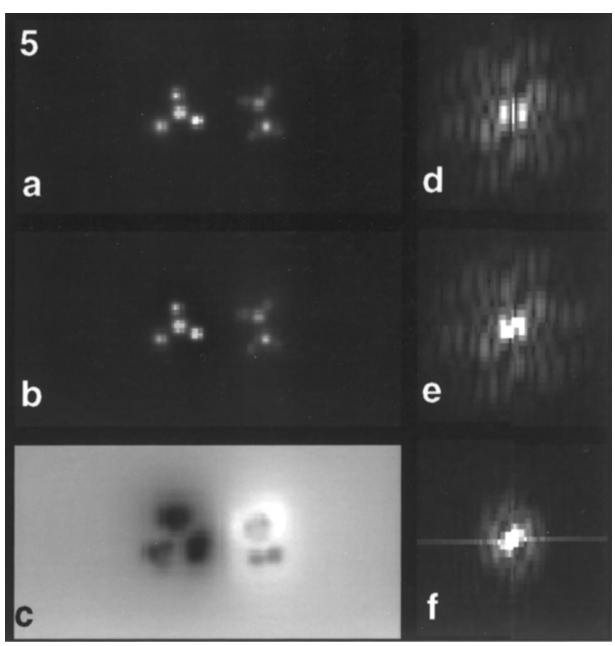

Fig. 5. As Fig. 2 but for an Asp–Lys ion pair. See text.

Sections of the three-dimensional potential maps for formamide and for the aspartic acid side chain, containing all the non-H atoms, are shown in Figs. 2 and 3. Because the nominal resolution is better than 1 Å, the images are dominated by the sharp high potential near the center of the atoms, and the maps for the free-atom and the molecular-orbital calculations appear nearly indistinguishable. Central sections of the three-dimensional Fourier transforms of each version of the Coulomb potential are shown beside the corresponding images. In both cases, there are noticeable differences between the free-atom and molecule calculations in the low-resolution range.

When the free-atom image is subtracted from the bonded-atom image, the low-resolution distortions in the potential of the bonded case become apparent in the difference images (Figs. 2*c* and 3*c*) and in the transforms (Figs. 2*f* and 3*f*). These distortions reflect expected features such as the shifts of electrons from the N atom towards the O atom, increasing the potential near the N atom while reducing it near the O atom.

Similar results for the hydrogen-bonded formamide pair are shown in Fig. 4. While a single formamide develops a substantial dipole across the molecule, a similar dipole exists for the pair. As a result, the charge distortions become localized to the distal ends and only small distortions appear in the region of the hydrogen bond itself. Again, differences are seen in the Fourier transforms, restricted to the low-resolution range.

The chemical bonding effects are even more dramatic in the case of the two residues that are expected to form an ion pair. Fig. 5 shows a section of the computed potential map that includes the heavy atoms in such a pair. The shift in the potential in the bonded calculation does in fact indicate a substantial ionization of the residues, with a change in the potential distribution that is substantially more visible than in the case of the formamide and aspartic acid models.

Fig. 6 presents the radially averaged transform amplitudes for these examples. Table 1 summarizes the *R* factors calculated in resolution zones for each of the four models. For these model compounds, bonding effects do indeed introduce very significant changes in the scattering-factor amplitudes, as seen in Fig. 6. For the covalently bonded compounds, the *R* factor is around 15% at the lowest resolutions computed in these examples, while for the ion pair it is over 40%. As expected, *R* factors become relatively small at higher scattering angles, especially beyond ~2.5 Å ($s = 0.4$).

Scattering factors tabulated for single neutral and ionized atoms (*International Tables for Crystallography*, 1974) indicate that the *R* factor drops below 1% beyond about $s = 0.5$. Since the *R* factors we observe are substantially larger, we checked to see that our methodology would replicate the previously tabulated data for single atoms. Fig. 7 shows plots of the scattering factors for neutral and ionized oxygen from the tables in

International Tables for Crystallography (1974) and from the *G92* calculations. The differences between the amplitudes for the neutral and ionized atoms does indeed become negligible at about the same scattering angle, and the *R* factor derived from the molecular-orbital calculations drops to around 1% at the same resolution as that from the tabulated data. However, by truncating the potential at 0.3 Å from the atom center we have reduced the long tail of the scattering factors, so that the *G92* curves fall off faster at higher angles than the more precisely calculated values. This effect does not significantly change the *R* factors within the resolution range considered here, though, and has no influence on the interpretation of the results.

Fig. 8 shows results from calculations of the RTP potential. RTP is non-planar, and the figure shows one section that includes two of the P atoms. These are clear in the potential map for both the neutral and charged models, along with one of the O atoms in the plane and part of the ribose (Figs. 8*a* and *b*). With a charge of −4, there is a strong halo of low potential around the phosphate chain, as expected (Fig. 8*b*). Images that are more relevant to the electron crystallography context are obtained by restricting the resolution to around 3 Å. When the resolution is truncated by a low-pass filter (Figs. 8*c* and *d*), the two phosphates in this plane and the ribose are seen as well separated blobs, although the individual atoms are not resolved. With a charge of −4, the negative halo causes a substantial loss of potential of the P atoms, nearly overwhelming the image of the phosphates. The low-frequency halo can, however, be nearly eliminated by then applying a high-pass filter to

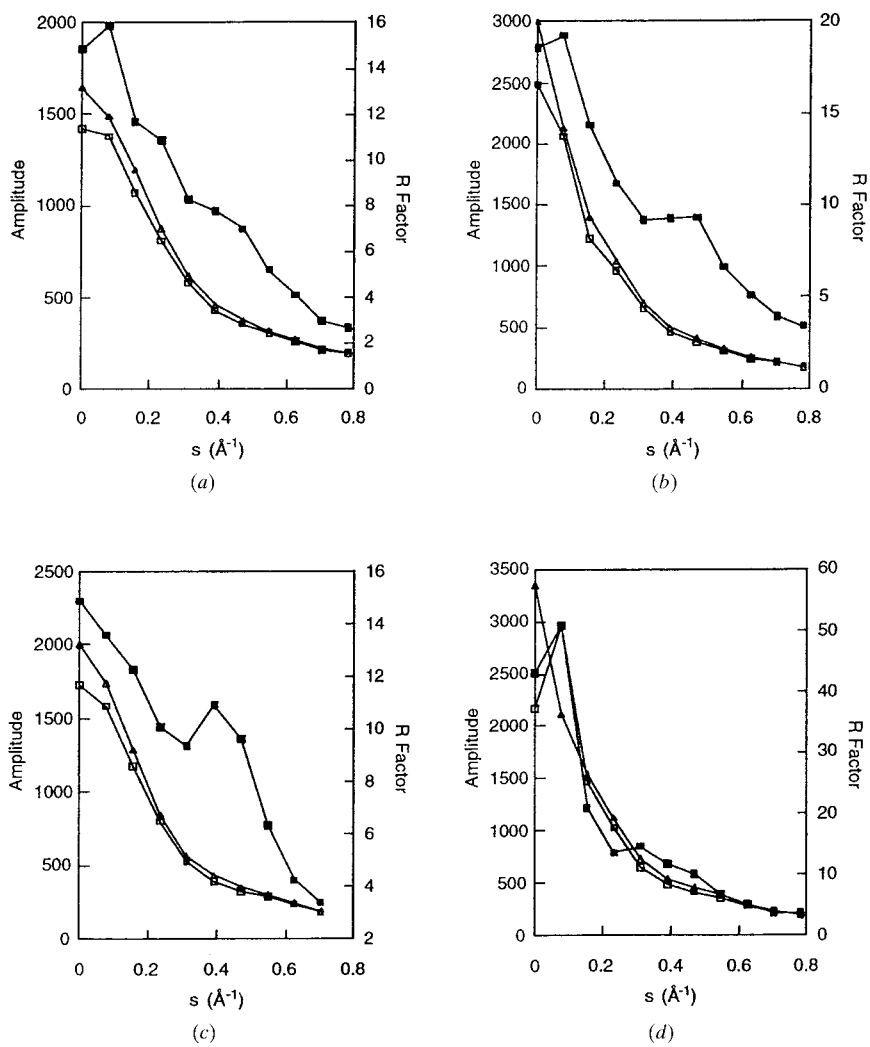

Fig. 6. Plots of radially averaged Fourier transform amplitudes and *R* factors for calculated models of (*a*) formamide (peptide bond), (*b*) pair of formamides (hydrogen-bonded pair), (*c*) aspartic acid side chain, (*d*) Asp–Lys ion pair. Amplitudes (arbitrary units) are averages within resolution zones, and *R* factors are calculated as in equation (3). Triangles: free-atom calculation; open squares: molecular calculation; filled squares: *R* factor.

Table 1. *R factors for the four models as a function of resolution*

$1/s$ (Å)	Formamide	Hydrogen-bonded pair	Aspartic acid	Ion pair
∞	14.84	18.56	14.88	42.94
12.8	15.84	19.25	13.55	50.90
6.40	11.67	14.35	12.23	20.79
4.23	10.86	11.17	10.09	13.54
3.20	8.24	9.21	9.36	14.62
2.56	7.75	9.29	10.90	11.77
2.13	6.98	9.34	9.62	10.08
1.83	5.21	6.64	6.34	6.85
1.60	4.11	5.11	4.26	4.99
1.42	2.94	3.97	3.40	3.76
1.28	2.66	3.42	3.03	3.61

exclude the very low resolution range. Figs. 8(*e*) and (*f*) show images calculated with data in the range 6–3 Å. Here the two images exhibit very similar contrast for all parts of the molecule.

4. Discussion

The modulation of electron scattering factors that results from redistribution of valence electrons in bonded materials is, for better or worse, significant just in the range of resolution within which data are most accessible in present studies of protein structure by electron crystallography. Once an incoming electron has penetrated within the valence shell of an atom, the potential it sees will be virtually the same as that of an isolated neutral atom. An electron passing through or outside the radius of the valence electrons, on the other hand, will see a potential that may be significantly different from that of an isolated atom. In the extreme case of an ionized atom, the potential must be inversely proportional to the radius at large distance, while for a neutral atom it is zero. Thus, for example, the scattering factor for any negatively charged atom must be negative over some range of (small) scattering angles. Among atoms found in proteins, negatively charged O atoms are the most extreme case; the scattering factor for singly ionized O atoms is negative for angles corresponding to spacings larger than about 5 Å. Beyond about 2 Å, however, the scattering factors for neutral and ionized O atoms are virtually indistinguishable. The perturbations of the other atomic potential functions that result from chemical bonding are expected to have similar, though less dramatic, effects within the same resolution range.

We find that the charge polarization within an amide bond, even when participating in formation of a hydrogen bond, can produce changes in scattering amplitudes large enough to add well over 10% to *R* factors in the resolution range up to around 5 Å. Even at a resolution of 2.5 Å, the chemical bonding effect can increase the *R* factor by about 5% if the molecular structure factors are calculated with independent-atom scattering factors. A similar magnitude of effect is found in the aspartic acid side chain, while substantially larger effects occur with an ion pair.

Electron diffraction data from proteins often show a larger error than X-ray data, with $R_{\rm merge}$ values typically 15–25% (Wang & Kühlbrandt, 1992; Kimura *et al.*, 1997; Nogales *et al.*, 1998). This error certainly contributes to the higher computed *R* factors seen in electron crystallography, but especially at low resolution this contribution may be less than that from chemical

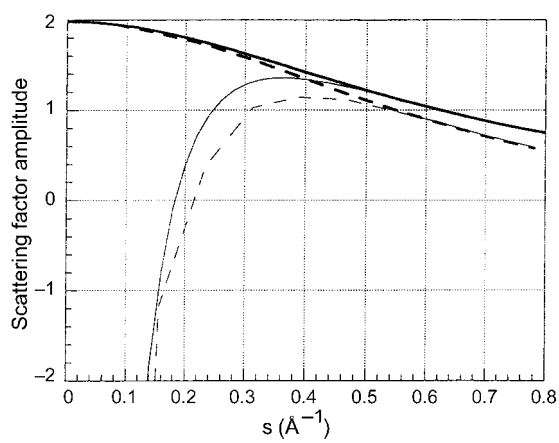

Fig. 7. Comparison of tabulated and calculated scattering factors for neutral (heavy lines) and ionized (thin lines) oxygen. Solid lines: data from *International Tables for Crystallography* (1974) plotted within the resolution range of interest. Dashed lines: radially averaged amplitudes from the three-dimensional Fourier transform of the *G92* calculation of single oxygen potentials, scaled to the tabulated data for the neutral atom at $s = 0$.

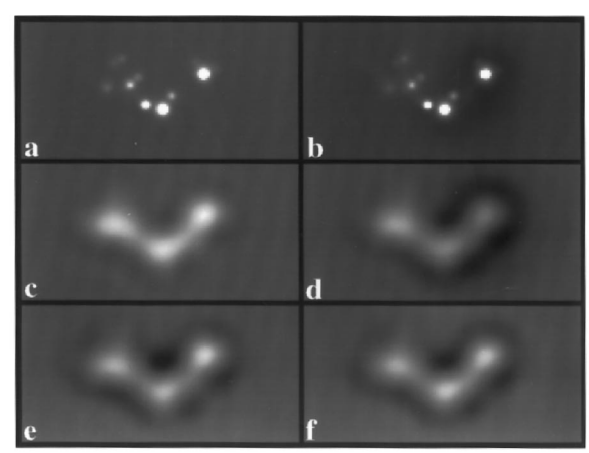

Fig. 8. Potential maps calculated for ribose triphosphate oriented as in Fig. 1. One section containing two of the P atoms is shown from the three-dimensional map. Calculations were performed for a neutral molecule [(*a*), (*c*), (*e*)] and for a net charge of −4 [(*b*), (*d*), (*f*)]. (*a*) and (*b*) show the potential computed with data in the full resolution range (24.8–0.4 Å). In (*c*) and (*d*), the resolution was truncated by imposing a low-pass filter with a cosine-function rolloff between 4 Å and 2.5 Å. In (*e*) and (*f*), the same low-pass filter was used, and the low-resolution range was eliminated by a high-pass filter with rolloff from 8–5 Å.

bonding effects. Thus it appears that the bonding effects are significant enough that they could seriously interfere with interpretation and refinement of density maps generated by electron crystallography.

The RTP model shows an interesting, though extreme, example of the effect of ionization. Near neutral pH, the phosphates may have a net charge of −4, with the resultant probability of producing a dramatic effect on the image. In any protein, of course, there will be neighboring positive charges to offset the effect, but Fig. 8 gives a clear example of how the image itself may be affected. With very high resolution, as in Fig. 8(*a*), very little difference appears between the images of the neutral and charged molecules. When the image resolution is limited to around 3 Å, as would generally be the case with protein crystals, the contrast for the negatively charged phosphates is decreased, in our model calculations by a factor of about two (Fig. 8*b*). This is the most dramatic example of the change in the image of the models calculated here, but similar effects would be present in every case. In the density map of tubulin, the only nucleotide-containing protein so far solved by electron crystallography (Nogales *et al.*, 1998), there was in fact identifiable density for the phosphates of GTP and GDP, but the density was well below that expected based on the usual scattering factors. Similar clear effects have been seen for amino acid side chains that form ion pairs in the tubulin map (unpublished data), in the LHC (Kühlbrandt *et al.*, 1994), and in the recent bR structure (Kimura *et al.*, 1997).

Some attempts have already been made to deal with problems that arise from the sensitivity of electron scattering to chemical bonding, although the success has been limited. In the electron crystallographic refinement study of bR (Grigorieff *et al.*, 1996), the electron scattering factors were refined by allowing the atomic occupancy to vary as a free parameter for each of 14 types of atom groups (*e.g.* C, CH, O^-, NH_3^+). Changes in the 'occupancy' as refinement progressed were then used to represent the increase or decrease in scattering strength of a fragment type. It was found that the trend in scattering potential, which decreases in the order of C, N and O for atomic electron scattering factors, is actually reversed when these atoms participate in chemical bonding in fragments such as the peptide backbone or a methylene group. The reason for this reversal was interpreted to be that the more incomplete valence shell of atomic carbon becomes filled when included in a molecule, so that shielding of the nucleus is increased. Oxygen, on the other hand, is better shielded than carbon in the atomic case, but the shielding is less efficient than carbon when the oxygen is part of a carbonyl fragment such as the peptide backbone or aspartic and glutamic acid side chains. The variation with charge was as expected: NH_3^+ and O^- showed higher and lower occupancies, respectively, than the corresponding neutral species.

The improvement in the *R* factor due to the refinement of the 'effective' electron scattering factor was found to be small, however (Grigorieff *et al.*, 1996). The small improvement may be due to the fact that adjusting the apparent occupancy, to scale up or scale down the strength of the shielded Coulomb potential, fails to capture the introduction of a long-range character when partial (or full) ionic species are present, including the large dipoles associated with ion pairs and the peptide backbone. The partial occupancy model also fails to capture the strong orientation dependence of the structure factor which is developed in such polar groups. Indeed, the dependence on orientation, for electron scattering, should make refinement more sensitive to orientation of polar groups than it can be in X-ray crystallography. While the chemical bonding effect is greatest for charged groups, the smaller effect arising from neutral covalent fragments may in the end affect the refinement even more since such groups represent a greater proportion of the protein structure.

The high-resolution electron crystallographic structure of LHC and the more recent refinement of the bR structure illustrate the limitations of atomic electron scattering factors in the case of charged residues. The weaker density of glutamic acid in the LHC structure (Kühlbrandt *et al.*, 1994) was probably a reflection of its negatively charged state, and the electron scattering factor for the carbonyl oxygen is probably better described by superoxide than by the neutral atom. In the most recent bR study (Kimura *et al.*, 1997), the higher-resolution data were used to generate an electrostatic potential map that, when compared with a map generated with low-resolution data, could be interpreted as indicating the charge states of most of the acidic residues of bR. While these studies reveal the limitations of atomic scattering factors applied to low-resolution data, they also illustrate a strength of electron microscopy relative to X-ray crystallography: the combination of low- and high-resolution data can be used to determine the charged state of amino acids important for biological function.

As pointed out by Kimura *et al.* (1997) and shown here in Fig. 8(*f*), when the low-resolution data are excluded in computing the potential, the effect of chemical bonding on the image can be largely suppressed. Similarly, if one used only the highest-resolution data in computing an overall *R* factor the value would be reduced. However, an even better approach is to use all of the available data and to incorporate perturbations of scattering factors when computing structure factors from an atomic model.

It is quite clear that better use of the low-resolution data for refinement can be realised by accounting for chemical bonding effects. In principle, the most accurate approach would be the calculation of structure factors based on *ab initio* evaluation of the molecular electrostatic potential (MEP) for the entire protein. However,

MEP calculation for a typical protein is a large computational task, and during the course of refinement the MEP might have to be evaluated many times before convergence is reached. Advances in linear scaling evaluation of the rate-limiting three-center one-electron integrals of the MEP and steady advances in computer hardware and parallelization should eventually allow calculations of this magnitude, but for the near future this is an intractable computational task.

Another approach is to modify the atomic form factors themselves to account for molecular bonding, *e.g.* by systematic development of molecular form factors based on a complete database of molecular fragments. Such an approach would be readily possible using the current and standard electronic structure algorithms and hardware. A database of the MEP of all relevant protein fragments, represented as a suitable expansion in a spherical harmonic basis, could be developed and used in place of the electron form factors, or as small perturbations on the usual factors, when defining model structure factors. The molecular form factors could be developed not only for covalently bound fragments but for fragments based on hydrogen-bonding geometries or ion pairs, which would account for important electron density at the boundaries between different portions of the backbone and/or side chain. This type of development is qualitatively similar to the proposed MEDLA (molecular electron density lego assembler) approach for constructing biopolymer electron density maps based on amino acid fragments (Walker & Mezey, 1994). The MEDLA approach matches fragment shapes and performs a simple average of the electron density between fragments at the adjacent boundary. Calculations of perturbations to the normal potentials would have the advantages of avoiding problems at fragment boundaries and of explicitly providing the contribution of bonding effects to the structure interpretation.

This work has been supported by National Institutes of Health grant GM51487 (KHD and RMG), the Air Force Office of Sponsored Research (FQ8671-9601129; THG) and by the Office of Health and Environmental Research, US Department of Energy, under contract DE-AC03-76F00098. We thank P. Walian and B. Jap for informative discussions, and Y. Fujiyoshi for a preprint of the high-resolution study of bR.

References

Anstis, G. R., Lynch, D. F., Moodie, A. F. & O'Keefe, M. A. (1973). *Acta Cryst.* A**29**, 138–147.

Branden, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

Doyle, P. A. & Cowley, J. M. (1974). In *International Tables for Crystallography*, Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

Doyle, P. A. & Turner, P. S. (1968). *Acta Cryst.* A**24**, 390–397.

Fox, A. G. (1993). *Philos. Mag. Lett.* **68**, 29–37.

Frisch, M. J., Trucks, G. W., Head-Gordon, M., Gill, P. M. W., Wong, M. W., Foresman, J. B., Johnson, B. G., Schlegel, H. B., Robb, M. A., Replogle, E. S., Gomperts, R., Andres, J. L., Raghavachari, K., Binkley, J. S., Gonzalez, C., Martin, R. L., Fox, D. J., Defrees, D. J., Baker, J., Stewart, J. J. P. & Pople, J. A. (1992). *Gaussian92*, revision *A*. Gaussian Inc., Pittsburgh, PA, USA.

Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. (1996). *J. Mol. Biol.* **259**, 393–421.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckman, E. & Downing, K. H. (1990). *J. Mol. Biol.* **213**, 899–929.

International Tables for Crystallography (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T. & Fujiyoshi, Y. (1997). *Nature (London)*, **389**, 206–211.

Koritsanszky, T., Flaig, R., Zobel, D., Krane, H.-G., Morgenroth, W. & Luger, P. (1997). *Science*, **279**, 356–358.

Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. (1994). *Nature (London)*, **367**, 614–621.

Nogales, E., Wolf, S. G. & Downing, K. H. (1998). *Nature (London)*, **391**, 199–203.

Spence, J. C. H. (1993). *Acta Cryst.* A**49**, 231–260.

Su, Z. & Coppens, P. (1992). *Acta Cryst.* A**48**, 188–197.

Vainshtein, B. K. (1964). *Structure Analysis by Electron Diffraction.* New York: MacMillian.

Walker, P. D. & Mezey, P. G. (1994). *J. Am. Chem. Soc.* **116**, 12022–12032.

Wang, D. N. & Kühlbrandt, W. (1992). *Biophys. J.* **61**, 287–297.